*Proc. Natl. Acad. Sci. USA*
Vol. 91, pp. 4579–4583, May 1994
Biochemistry

# Similarity between a ubiquitous promoter element in an ancient eukaryote and mammalian initiator elements

(*Trichomonas vaginalis*/gene regulation)

DORIS V. K. QUON*, MARIA G. DELGADILLO*, ANITA KHACHI*†, STEPHEN T. SMALE*†‡, AND PATRICIA J. JOHNSON*‡§

*Department of Microbiology and Immunology, †Howard Hughes Medical Institute, and ‡Molecular Biology Institute, University of California, Los Angeles, School of Medicine, Los Angeles, CA 90024-1747

*Communicated by Elizabeth F. Neufeld, January 25, 1994 (received for review December 13, 1993)*

ABSTRACT **To identify regulatory elements that play a role in transcription initiation in ancient eukaryotes, we have analyzed the upstream regions of protein-coding genes from *Trichomonas vaginalis*, one of the most ancient eukaryotes studied to date. Characterization of seven protein-coding genes from this protist invariably revealed the presence of a highly conserved DNA sequence motif immediately upstream of the coding region. This 13-nt motif was shown to surround and contain precise sites for transcription initiation. No typical TATA boxes, positioned at 25–30 nt upstream of the transcription start sites of these genes, were found. The start-site regions from all seven *T. vaginalis* genes impart strong specific initiation of transcription in a mammalian *in vitro* transcription assay. This consensus promoter element in an ancient eukaryote is similar, both structurally and functionally, to initiator elements found in promoters of higher eukaryotes.**

Transcriptional regulation of protein-coding genes has been studied extensively in higher eukaryotes, leading to the identification of conserved DNA sequence elements responsible for the accurate initiation of mRNA synthesis by RNA polymerase II. At least two types of core promoter elements exist, TATA boxes and initiator (Inr) elements. The TATA box, a conserved element found in the promoter regions of many genes, contains the sequence motif TATAAA (1, 2). This TATA motif is usually located 25–30 bp 5′ of the site of transcription initiation and plays a critical role in start-site selection and recruitment of the general transcription machinery to the promoter (1, 3). A second less-abundant class of promoters lack a TATA box but contain an element called an Inr, which surrounds the RNA start site and appears to carry out the same functions as TATA (4–8). The remaining core promoters contain both TATA and Inr elements or neither.

The evolutionary origins of the structural components of eukaryotic core promoters are not known. TATA-like promoter sequences have been reported in archaebacteria (9). In addition, the location of the TATA box relative to the transcription start site is similar to the location of a critical element found at −35 in eubacterial promoters, but it is not clear whether this similarity is relevant to the origin of the TATA box. The DNA-binding domain of the TATA-binding protein has sequence similarity with prokaryotic $\sigma$ factors; however, the similarity is restricted to the region of the $\sigma$ factors thought to interact with the −10 element of bacterial promoters (3, 4).

One reason for our lack of knowledge about the evolutionary origin of eukaryotic transcriptional control elements and transcription initiation mechanisms is that, to date, the study of eukaryotic transcription has been confined largely to animals, plants, and fungi. As illustrated by the phylogenetic tree shown in Fig. 1, these organisms encompass a relatively limited range of eukaryotic evolution. Until recently, the biological diversity represented by protist lineages has received little attention by molecular biologists. The only protists for which RNA polymerase II promoters have been clearly defined are advanced protists: the slime mold, *Dictyostelium* (12, 13), and the amoeba, *Acanthamoeba* (14) (Fig. 1). Genes that have been analyzed from these organisms appear to contain TATA boxes ≈30 bp from the transcription start site (13–16). Furthermore, TATA-binding protein genes similar to those isolated from higher eukaryotes have been identified in *Acanthamoeba* and *Plasmodium* (14, 17).

We recently have begun an analysis of protein-coding genes from a more primitive protist, *Trichomonas vaginalis*. This flagellated protist is a sexually transmitted human parasite that is a common cause of vaginitis. Molecular phylogenetic studies using 18S and 28S rRNA (11, 18, 19) indicate that this single-celled organism is among the most ancient eukaryotes studied thus far. *Trichomonas* diverged from the main line of eukaryotic descent prior to the divergence of most other protist lineages (Fig. 1). Consistent with an early divergence, trichomonads lack two hallmark organelles of eukaryotic cells: mitochondria and peroxisomes (for review, see refs. 20 and 21). Given its phylogenetic position, *T. vaginalis* provides a system for identifying molecular properties of transcription in the earliest eukaryotic organisms.

In this communication, we report the presence of a highly conserved DNA sequence motif found upstream of all *T. vaginalis* protein-coding genes reported to date.¶ The 5′ ends of mapped mRNAs invariably begin within this conserved sequence. This motif is also capable of imparting specific transcription initiation in mammalian *in vitro* transcription assays. These data indicate that the conserved element is a core promoter that acts to direct the initiation site of transcription.

## MATERIALS AND METHODS

***T. vaginalis* Cultures.** *T. vaginalis* cultures (ATCC 30001) were grown in Diamond's medium, supplemented with 10% (vol/vol) horse serum and iron as described (22).

**PCR Amplification.** P-glycoprotein (Pgp1), β-tubulin (β-Tub), 70-kDa cytosolic heat shock protein (cHSP70), and the α subunit of succinyl CoA synthetase (α-SCSB) were isolated

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Inr, initiator; Tub, tubulin; cHSP70, 70-kDa cytosolic heat shock protein; α-SCSB, α subunit of succinyl CoA synthetase; Pgp, P-glycoprotein; Fd, ferredoxin; β-SCS, β-subunit of succinyl CoA synthetase.

§To whom reprint requests should be addressed.

¶The sequences reported in this paper have been deposited in the GenBank data base (accession nos. αSCS, V07779; αTub, V07780; BSCS, V07781; βTub, V07782; CHSP70, V07783; FD, V07784; Pgp1, V07785).

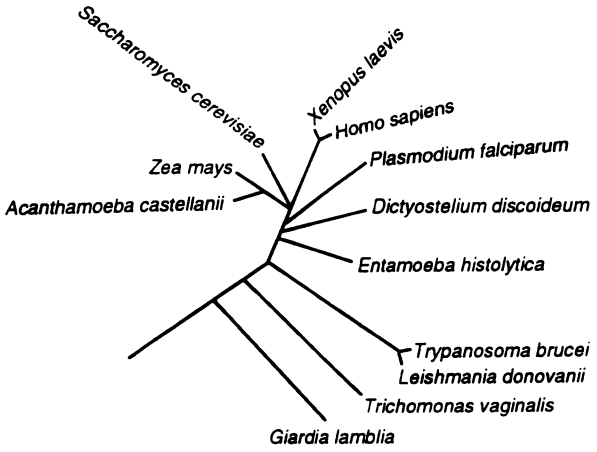


FIG. 1. Eukaryotic evolutionary tree inferred from 18S rRNA-like sequence similarities. Alignments take into account both primary and secondary structural features, using the distance matrix methods of Fitch and Margoliash (10). The earliest diverging lineages, represented by *Giardia lamblia* and *T. vaginalis*, are amitochondrial protists. The protists are a large diverse group that diverged long before the relatively recent radiation of the major lineages leading to fungi, plants, and animals (modified from ref. 11).

using PCRs. PCRs were performed using 1 μg or 100 ng of genomic DNA in 50 mM KCl/10 mM Tris·HCl/1.5 mM $MgCl_2$/250 pmol of primer/all four dNTPs (each at 200 mM)/2.5 units of *Taq* polymerase (Cetus). PCR cycling temperatures used for Pgp1, cHSP70, and β-Tub were 95°C, 45°C, and 72°C. Cycling temperatures for α-SCSB were 95°C, 52°C, and 72°C. Degenerate oligonucleotide primers were as follows: Pgp, primer 1, (T/A)C(C/A/T)GG(T/C)T(C/G)(C/A)GG(T/C)AAGTC(C/A/T)ACA, corresponding to the peptide (S/T)-G(S/C)GKST, and primer 2, CT(C/T)TC(C/A/T)GG(T/C)GG(T/C)CA(A/G)AAGCA(A/G), inversely complementary to LSGGQKQ; β-Tub, primer 1, TGGGC(C/T)A-AGGG(C/T)CA(C/T)TA(C/T)ACA, corresponding to WAKGHYT, and primer 2, GC(G/A)AAGGGTCC(G/A)GT(T/C)GA(G/A)TTG, inversely complementary to RF-PGQLN; cHSP70, primer 1, GG(C/T)AT(C/T)GA(C/T)CT(C/T)GG(C/T)ACAACATA, corresponding to GIDL-GTTY, and primer 2, GT(C/T)CG(G/A)TGTTTCCT(G/A)CG(G/A)CC, inversely complementary to QATKDAG; α-SCSB, primer 1, GT(C/T)ACACCAGG(T/C)AAGGG(C/T)GG, corresponding to VTPGKGG, and primer 2, GC(C/T)TC(G/A)TATGT(G/A)AGTGT(G/A)CC, inversely complementary to GTLTYEA. Final PCR products were cloned into the *Eco*RV site of pBluescript and sequenced by the Sanger dideoxynucleotide method (United States Biochemical).

**Isolation and Cloning of Genes.** The genes for ferredoxin (Fd) and the β subunit of succinyl CoA synthetase (β-SCS) were cloned as described (23, 24). The α-Tub gene was obtained by screening a *T. vaginalis* genomic library in λEMBL3 (23) with a $^{32}$P-random-primed-labeled α-Tub cDNA probe (pTbαT-c1) from *Trypanosoma brucei* (25). Filters were washed to a final stringency of 3× standard saline citrate at 65°C. PCR products, Pgp, β-Tub, cHSP70, and α-SCSB, were used to screen either *T. vaginalis* λEMBL3 (Pgp, cHSP70, and β-Tub) or λZAP genomic libraries (α-SCSB). The λEMBL3 library was made as described (23). The λZAP (Stratagene) libraries were made either by partially digesting total genomic DNA with *Alu* I, blunting the ends with the Klenow fragment of DNA polymerase I, and adding *Not* I–*Eco*RI linkers or by completely digesting the DNA with *Eco*RI. The fragments were ligated into λZAP arms (Stratagene), packaged, and plated as recommended by the manufacturer. Libraries were screened with either a random-primed $^{32}$P-labeled or digoxigenin-labeled probe (Boehringer Mannheim). The clones isolated using digoxigenin-labeled probes were identified by alkaline phosphatase color reactions.

**RNA Isolation.** RNA was isolated by a LiCl/urea method. Approximately $10^{10}$ cells were grown and harvested. The cells were washed twice with 1× phosphate-buffered saline (PBS) and resuspended in 5 ml of 1× PBS. Forty milliliters of ice-cold 3 M LiCl/6 M urea was added rapidly and the cells were homogenized on ice in a Sorvall Omnimixer (model 17150) at the highest speed for 1 min. The RNA was precipitated overnight on ice. The RNA was pelleted in an SW28 rotor at 25,000 rpm at 4°C for 1 h. The RNA was resuspended, proteinase K-treated (0.2 mg/ml), and extracted with phenol/chloroform. The RNA was resuspended and stored in ethanol.

**Primer-Extension Analysis.** Primers were made to sequences ≈125 nt from the predicted 5′ end of the mRNAs. Primers used are as follows: Fd, CGTGCTTGCAGATG-CACTT; β-SCS, AGCCTCTTCTGGGGAACG; α-Tub, TGGGAGCTGGCCATCTGG; Pgp, TACCCATCTGT-GATCGAA; α-SCSB, CTTGGATAACAACTCTT; cHSP70, GGTGTTGTTCTGTTACCC; and β-Tub, GTGG-TATGAACCTGTTGG. Primers were 5′-end-labeled with [γ-$^{32}$P]ATP and T4 polynucleotide kinase (New England Biolabs). Total *T. vaginalis* RNA was hybridized with the $^{32}$P-labeled primers in 250 mM KCl/5 mM Tris·HCl, pH 7.5/1 mM EDTA at 37°C for 90 min. Reverse transcriptase reactions were performed at 37°C using Moloney murine leukemia virus reverse transcriptase (BRL) with all four dNTPs (each at 50 μM) as described (26). Sequencing was performed using the Sanger dideoxynucleotide method (United States Biochemical).

***In Vitro* Transcription in HeLa Cells.** Oligonucleotides corresponding to the *T. vaginalis* Inr elements or the control elements terminal deoxynucleotidyltransferase, R2549, R2533, and dihydrofolate reductase were synthesized and inserted into the *Bam*HI and *Sac* I sites in plasmid III (27), containing multiple Sp1 binding sites and a very weak TATA box (28). Nuclear extracts were prepared as described (29). *In vitro* transcription reactions were performed with 100 μg of unfractionated HeLa cell nuclear extracts and 300 ng of template DNA. RNA was analyzed by primer-extension analysis with a $^{32}$P-end-labeled SP6 promoter primer (28). cDNA products were analyzed on an 8% denaturing polyacrylamide gel.

## RESULTS

**Characterization of Sequences Upstream of *T. vaginalis* Protein-Coding Genes.** As a step toward identifying DNA sequence motifs that play a role in gene expression in trichomonads, we have isolated and characterized seven *T. vaginalis* protein coding genes. The genes encode Fd (23), β-SCS (24), α-SCSB, α-Tub, cHSP70, Pgp1, and β-Tub. These proteins represent different classes of molecules (e.g., cytoskeletal structural proteins vs. enzymes) that are found in different locations in the cell (e.g., cytosolic vs. organellar). Analysis of the DNA sequence upstream of the seven genes identified a conserved pyrimidine-rich sequence motif immediately 5′ of the ATG initiation codon (Fig. 2). The sequence conservation of this 13-nt motif is remarkable. In all seven genes, 3 nt are absolutely conserved, the adenosine at position +1, the cytidine at position −1, and the thymidine at position +3 (Fig. 2). In addition, a specific nucleotide is conserved at positions −2, +7, +8, and +11 in six of the seven genes. This conservation results in the strong consensus sequence, $TCA^{+1}YTWYTCATTA$. It is notable that no appropriately spaced TATA boxes or other cis elements commonly associated with RNA polymerase II promoters were identified.

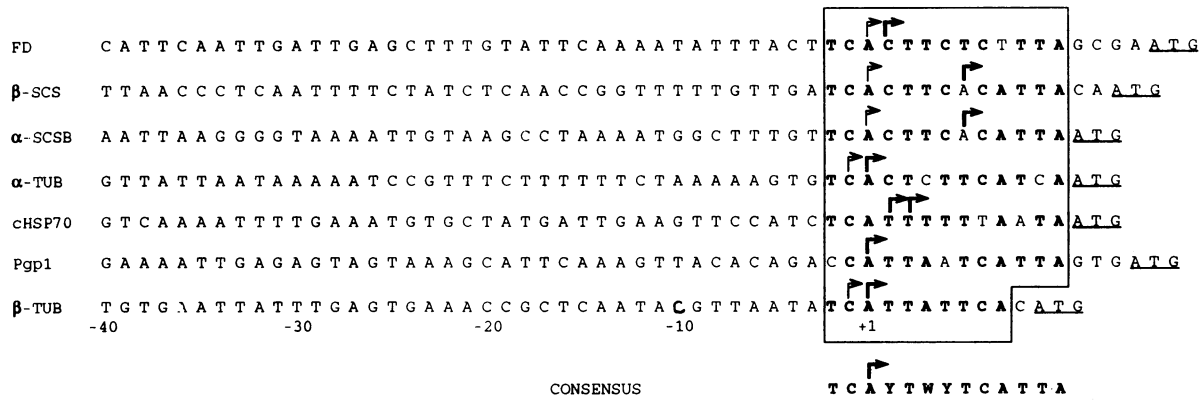


FIG. 2. Conserved nucleotide sequence surrounding the transcription initiation sites of seven *T. vaginalis* protein coding genes. The 5′ sequences immediately flanking the ATG translational start codon of genes encoding the following proteins are shown: Fd (23), β-SCSB (24), α-SCSB, α-Tub, cHSP70, Pgp1, and β-Tub. The conserved motif is boxed and a consensus sequence is listed below. Note that this motif is identical for the α-SCSB and β-SCS genes. Arrows mark the site of transcription initiation, as mapped using primer-extension assays (see Fig. 3). Boldface arrows denote strong initiation sites; lightface arrows denote weaker sites. Nucleotides have been numbered by designating the adenosine that is most frequently used as the initiation site +1.

**Identification of Transcriptional Start Sites.** To address the biological relevance of this motif, we employed primer-extension analyses to map the sites of transcription initiation for each gene. Representative data are shown in Fig. 3 for the α-Tub and β-Tub genes and the cHSP70 gene. The 5′ ends of the mRNA for all seven genes are indicated by arrows above the sequences in Fig. 2. Without exception, the predominant transcripts began within the conserved 13-nt motif, either at an invariant adenosine or ±1 nt from this residue. The 5′ end of mRNAs that do not map to an adenosine map to either an adjacent cytidine or thymidine. Minor start sites were also observed in five genes (indicated by smaller arrows). All primer-extension experiments have been repeated several times with different mRNA preparations. Nonetheless, it should be noted that reverse transcriptase may terminate a few nucleotides from the end of the template and thus the mapped sites may, in fact, not represent the exact 5′ end of the transcripts. For the Fd and β-SCS mRNAs, the primer-extension results were confirmed by S1 nuclease analyses, which also localized the 5′ ends to nucleotides within the conserved motif (23, 24). These mRNAs have been shown to contain a 5′ cap structure (data not shown). Furthermore, sequence comparison of the 5′ ends of the cDNA and genomic clones for Fd and β-SCS shows colinearity (23, 24), eliminating the possibility that the 5′ ends undergo trans-splicing as occurs in kinetoplastid protists (31–33). Thus, these observations provide strong evidence that the 5′ ends mapped for the mRNAs correspond to the site of transcription initiation and not a 5′ processing site.

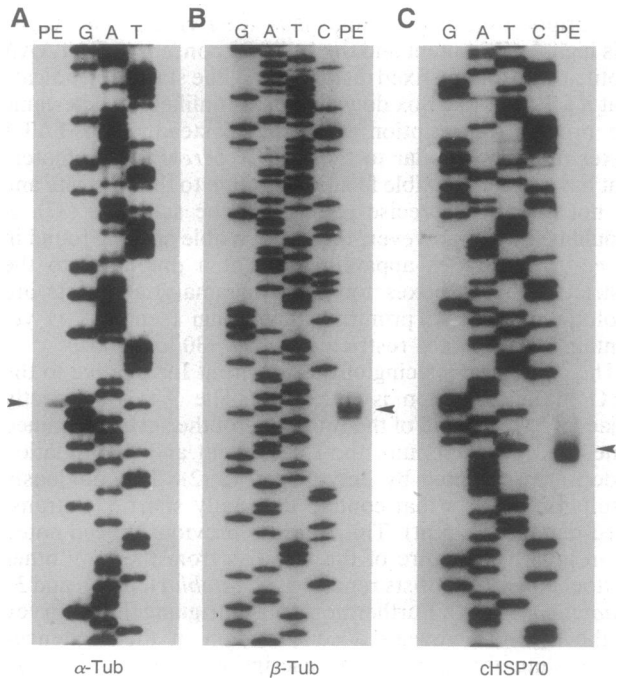


FIG. 3. Primer-extension analysis to determine transcriptional initiation sites in *T. vaginalis* genes. $^{32}$P-end-labeled primers were hybridized to total RNA and primer-extension reactions were performed with Moloney murine leukemia virus reverse transcriptase as described (30). (*A*) Primer-extension reaction using primers to the α-Tub gene. The primer-extension product is shown in lane PE. The sequence ladder reads G, A, and T. (*B*) Same as *A* except primers correspond to the β-Tub gene and the sequence ladder is G, A, T, and C. The primer-extension product is in lane PE. (*C*) Same as *B* except primers correspond to cHSP70.

Given the consistent location of the transcription start sites within the 13-bp DNA sequence motif and its remarkable degree of sequence conservation, we propose that this motif acts as an Inr element, by recruiting the RNA polymerase and associated transcription factors necessary for accurate transcription initiation.

***In Vitro* Transcription in HeLa Cell Nuclear Extracts.** The similarity between the *T. vaginalis* Inr element and those of higher eukaryotes led us to test the ability of the *T. vaginalis* Inr elements to direct the initiation of RNA synthesis in a mammalian *in vitro* transcription assay. Inr elements from *T. vaginalis* (from positions −6 to +11 relative to the adenosine at +1) were inserted into a plasmid containing multiple binding sites for transcription factor Sp1 but lacking a strong TATA box (27). These plasmids were then incubated with HeLa cell nuclear extracts and NTPs. All Inr elements were inserted into the same location in the plasmid, so that transcription initiation from the predicted site would produce the same size cDNA products upon analysis of the transcripts by primer extension.

Each of the seven *T. vaginalis* sequences imparted strong Inr activity in the heterologous system, as indicated by the fact that the presence of a *T. vaginalis* Inr element greatly enhanced the efficiency of transcription originating from a site located within the element (Fig. 4). Each of the trichomonad Inr elements was almost as strong as the well-characterized mammalian terminal deoxynucleotidyltransferase Inr (lane 1) (7, 28, 29). The transcripts produced *in vitro* initiated specifically at the predicted adenosine (indicated by the arrow in Fig. 4) with the exception of those directed by the cHSP70 Inr, which appeared to initiate at both the

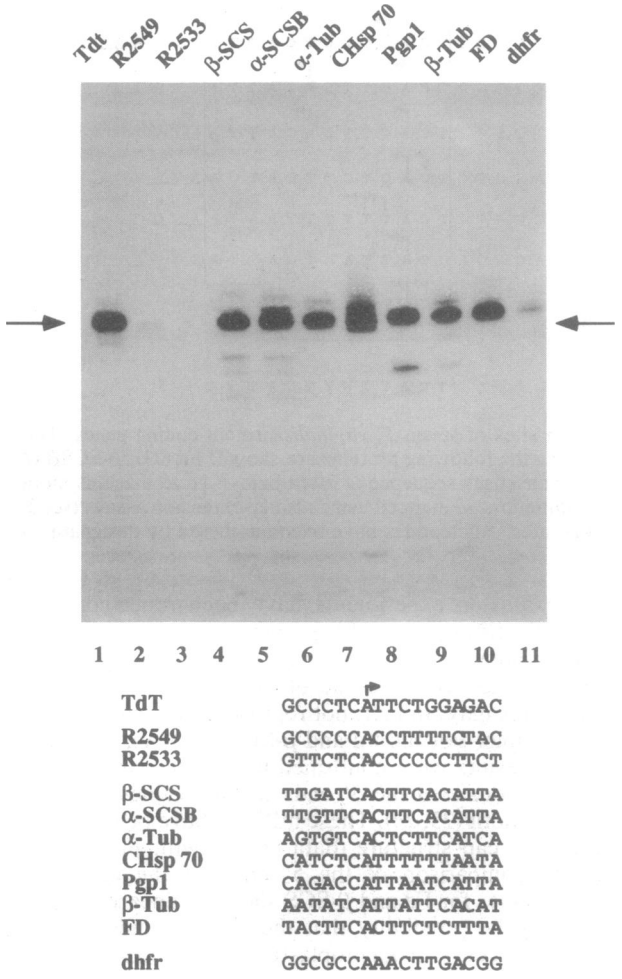


FIG. 4. *In vitro* transcription in HeLa nuclear extracts using *T. vaginalis* Inr elements. HeLa cell nuclear extracts were prepared as described (29). *In vitro* transcription assays were performed and analyzed by primer extension with an SP6 promoter primer (28). The expected products of 79 nt are indicated by the arrow. The reactions contained plasmids with the terminal deoxynucleotidyltransferase (TdT) Inr (lane 1), the R2549 and R2533 Inr elements (lanes 2 and 3), the seven *T. vaginalis* Inr elements (lanes 4–10, specific Inr elements indicated above each lane), and the mammalian dihydrofolate reductase (dhfr) Inr (lane 11) (34).

adenosine and the adjacent thymidine. The sequence specificity of this assay is clearly demonstrated by the lack of initiation observed for two negative control elements, R2549 and R2533 (lanes 2 and 3). These elements contain the sequence CA (−1 and +1) surrounded by an array of pyrimidines that do not create functional Inr elements (35). It is also noteworthy that the *T. vaginalis* sequences are stronger Inr elements in the mammalian extract than the basal Inr derived from a mammalian dihydrofolate reductase promoter (34, 35) (lane 11).

The ability of *T. vaginalis* Inr elements to direct accurate transcription initiation in HeLa nuclear extracts and the observation that the 5′ ends of *T. vaginalis* mRNAs invariably begin within this conserved motif provide strong evidence that this motif functions as an Inr *in vivo*. Direct functional testing of these elements must await the development of transformation and *in vitro* transcription assays for *Trichomonas*.

## DISCUSSION

To date no RNA polymerase II promoter elements have been clearly defined in ancient protists. To this end, we have isolated and compared the sequences immediately upstream of seven *T. vaginalis* protein coding genes. These sequences represent all those currently available for comparison. The upstream sequences of all seven genes contain the consensus sequence, TCAYTWYTCATTA. Furthermore, the 5′ ends of the mRNAs encoded by the genes invariably map within this conserved motif and the motif is capable of directing specific transcription initiation in HeLa cell nuclear extracts. These data strongly suggest a role for this element in transcription initiation, similar to that previously defined for Inr elements in higher eukaryotic genes.

The conservation of the element 5′ of *T. vaginalis* protein coding genes is striking. The consensus start site adenosine is conserved in all seven genes, as well as the thymidine at −1 and the cytidine at +3. In addition, six of the seven genes contain the same nucleotides in four other positions (Fig. 2). The sequences of Inr elements in higher eukaryotes are not as highly conserved as those identified in *T. vaginalis*. However, recent functional studies of a large series of random and mutant Inr elements have established a loose consensus sequence for mammalian Inr activity as $YCA^{+1}NTYY$, where Y is a pyrimidine (35). Furthermore, a survey of >500 higher eukaryotic genes revealed that cytidine at −1, adenosine at +1, and thymidine at +3 are the three most highly conserved nucleotides at the start-site regions of these genes (2). Interestingly, these are the nucleotides that are strictly conserved in the Inr elements of *T. vaginalis*.

Examination of the DNA sequences upstream of the Inr elements failed to reveal appropriately spaced TATA boxes or other cis elements frequently associated with eukaryotic RNA polymerase II promoters (1, 36). However, our current data do not preclude the presence of distal TATA elements. TATA activity can be imparted by numerous A+T-rich sequences and the sequence upstream of the *T. vaginalis* transcription start sites is unusually A+T-rich. Nonetheless, this high A+T content and the lack of a consensus TATAAA motif at −30 (or any fixed distance from the start site) indicate that if a TATA-like box does exist, it is unlikely to determine the precise transcription start site. Instead, such TATA boxes might be similar to those in *S. cerevisiae* promoters that have a more flexible location relative to the start site and do not direct the precise position of the start site (37). It should be noted, however, that this flexible spacing found in *S. cerevisiae* genes apparently is not a precursor to the restricted TATA boxes found in mammals, as plants are evolutionarily more primitive than fungi (see Fig. 1) yet contain TATA boxes restricted to the −30 location.

The conserved spacing of this apparent Inr relative to the ATG initiation codon is intriguing. The ATG is directly adjacent to the 3′ end of the Inr in four of the seven examined genes and, at maximum, the Inr element and the initiation codon are separated by only 4 nt (Fig. 2). This obviously results in mRNAs that contain unusually short 5′ untranslated regions (6–15 nt). This trait has previously been noted as an invariant feature of the mRNAs from the only other amitochondriate protists reported, *G. lamblia* (38–40) and *E. histolytica* (41–44). Furthermore, it is intriguing that a survey of the literature reveals a conservation of the sequences surrounding the apparent transcription initiation sites of protein-coding genes of *G. lamblia* (38–40) and *E. histolytica* (41–44). The 5′ end of the only four reported *Entamoeba* mRNAs maps within the sequence ATTCA. The 3′ end of this sequence is only 7–12 nt upstream of the ATG initiation codons of these genes. Interestingly, this pentamer contains an absolutely conserved TCA sequence, similar to that observed in the *T. vaginalis* Inr elements. The 5′ ends of the majority of reported *Giardia* mRNAs begin within a short oligo(T)-oligo(A) sequence that abuts the ATG initiation codon (38–40). Consistent with the findings reported here for

*T. vaginalis* genes, *E. histolytica* and *G. lamblia* genes also lack appropriately spaced TATA boxes (38–44). In fact, obvious TATA boxes have yet to be observed ≈30 nt upstream of the transcription start site of protein-coding genes found in any protist lineage (26, 45–49) prior to the lineage that gave rise to *Dictyostelium* (Fig. 1).

*T. vaginalis* is the most ancient eukaryote for which a highly conserved motif that appears to play a role in transcription has been reported. The conservation of this element in all protein-coding genes analyzed from *T. vaginalis* and its ability to direct transcription *in vitro* strongly suggest that the Inr is an essential element for basal gene transcription in trichomonads. Because the *T. vaginalis* genome is simpler and requires less regulatory complexity than is needed in higher eukaryotes, promoter strength may rely heavily on this element. It is interesting to note that several of the genes appear to contain tandem repeats of the sequence CANT within the 13-bp conserved motif. Perhaps the presence of more than one Inr element may create a stronger promoter. Direct analysis of the role this Inr element plays in gene transcription awaits the development of a transformation or an *in vitro* assay in *Trichomonas*.

The structural and functional similarity of *T. vaginalis* and mammalian Inr elements is intriguing. Given that early eukaryotes may be as ancient as eubacteria and archaebacteria (50), it is tempting to speculate that Inr elements arose early during evolution and were later supplemented by TATA elements. Identification of additional promoters in amitochondrial protists and other primitive eukaryotes will help address this possibility. Such studies should also lead to a better understanding of the evolution of eukaryotic promoter structure.

We thank Arnie Berk, Mike Carey, and Larry Simpson for critically reviewing the manuscript; Carol Lahti, Christine d'Oliveira, and Beatrice Schuck for cloning α-SCSB, α-Tub, and Pgp1 genes, respectively; and Jeff Quicho for helping us prepare figures. This work was supported by grants from the National Institutes of Health (AI 30537 and AI27857 to P.J.J.), a predoctoral training grant (GM-08042 to D.V.K.Q.), and a Burroughs-Wellcome New Investigator in Molecular Parasitology Award to P.J.J.

1. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
2. Bucher, P. (1990) *J. Mol. Biol.* **212,** 563–578.
3. Zawel, L. & Reinberg, D. (1992) *Curr. Opin. Cell Biol.* **4,** 488–495.
4. Zuber, P., Healy, J., Carter, H. L., Cutting, S., Moran, C. P. & Losick, P. (1989) *J. Mol. Biol.* **206,** 605–614.
5. Weis, L. & Reinberg, D. (1992) *FASEB J.* **6,** 3300–3309.
6. Smale, S. T. (1994) in *Transcription: Mechanisms and Regulation*, eds. Conaway, R. & Conaway, J. W. (Raven, New York), pp. 63–81.
7. Smale, S. T. & Baltimore, D. (1989) *Cell* **57,** 103–113.
8. Kollmar, R. & Farnham, P. J. (1993) *Proc. Soc. Exp. Biol. Med.* **203,** 127–139.
9. Reiter, W. D., Hudepohl, U. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9509–9513.
10. Fitch, W. M. & Margoliash, E. (1967) *Science* **155,** 279–284.
11. Sogin, M. L. (1989) *Am. Zool.* **29,** 487–499.
12. Kimmel, A. R. & Firtel, R. A. (1983) *Nucleic Acids Res.* **11,** 541–552.
13. Cohen, S. M., Knecht, D., Lodish, H. F. & Loomis, W. F. (1986) *EMBO J.* **5,** 3361–3366.
14. Wong, J. M., Liu, F. & Bateman, E. (1992) *Nucleic Acids Res.* **20,** 4817–4824.
15. Nellen, W. & Gallwitz, D. (1982) *J. Mol. Biol.* **159,** 1–18.
16. Nellen, W., Silan, C., Saur, U. & Firtel, R. A. (1986) *EMBO J.* **5,** 3367–3372.
17. McAndrew, M. B., Read, M., Sims, P. F. & Hyde, J. E. (1993) *Gene* **124,** 165–171.
18. Baroin, A., Perasso, R., Qu, L. H., Brugerolle, G., Bachellerie, J. P. & Adoutte, A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3474–3478.
19. Leipe, D. D., Genderson, J. H., Nerad, T. A. & Sogin, M. L. (1993) *Mol. Biochem. Parasitol.* **59,** 41–48.
20. Müller, M. (1988) *Annu. Rev. Microbiol.* **42,** 465–488.
21. Johnson, P. J., Lahti, C. J. & Bradley, P. J. (1993) *J. Parasitol.* **79,** 664–670.
22. Gorrell, T. E., Yarlett, N. & Müller, M. (1984) *Carlsberg Res. Commun.* **49,** 259–268.
23. Johnson, P. J., d'Oliveira, C. E., Gorrell, T. E. & Müller, M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6097–6101.
24. Lahti, C. J., d'Oliveira, C. E. & Johnson, P. J. (1992) *J. Bacteriol.* **174,** 6822–6830.
25. Thomashow, L. S., Milhausen, M., Rutter, W. J. & Agabian, N. (1983) *Cell* **32,** 35–43.
26. Zomerdijk, J. C. B. M., Ouellette, M., ten Asbroek, A. L., Kieft, R., Bommer, A. M., Clayton, C. E. & Borst, P. (1990) *EMBO J.* **9,** 2791–2801.
27. Smale, S. T., Schmidt, M. C., Berk, A. J. & Baltimore, D. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4509–4513.
28. Zenzie-Gregory, B., O'Shea-Greenfield, A. & Smale, S. T. (1992) *J. Biol. Chem.* **267,** 2823–2830.
29. O'Shea-Greenfield, A. & Smale, S. T. (1992) *J. Biol. Chem.* **267,** 1391–1402.
30. McKnight, S. L. & Kingsburg, R. (1982) *Science* **217,** 316–324.
31. Murphy, W. J., Watkins, K. P. & Agabian, N. (1986) *Cell* **47,** 517–525.
32. Sutton, R. E. & Boothroyd, J. C. (1986) *Cell* **47,** 527–535.
33. Laird, P. W., Zomerdijk, J. C. B. M., DeKorte, D. & Borst, P. (1987) *EMBO J.* **6,** 1055–1062.
34. Means, A. L. & Farnham, P. J. (1990) *Mol. Cell. Biol.* **10,** 653–661.
35. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. (1994) *Mol. Cell. Biol.* **14,** 116–117.
36. Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58,** 799–839.
37. Struhl, K. (1987) *Cell* **49,** 295–297.
38. Kirk-Mason, K. E., Turner, M. J. & Chakraborty, P. R. (1989) *Mol. Biochem. Parasitol.* **36,** 87–99.
39. Alonso, R. A. & Peattie, D. A. (1992) *Mol. Biochem. Parasitol.* **50,** 95–104.
40. Ey, P. L., Khanna, K., Manning, P. A. & Mayrhofer, G. (1993) *Mol. Biochem. Parasitol.* **58,** 247–257.
41. Edman, U., Meraz, M. A., Rausser, S., Agabian, N. & Meza, I. (1990) *J. Exp. Med.* **172,** 879–888.
42. Huber, M., Garfinkel, L., Mirelman, D., Revel, M. & Rozenblatt, S. (1987) *Mol. Biochem. Parasitol.* **24,** 227–235.
43. Huber, M., Garfinkel, L., Gitler, C., Mirelman, D., Revel, M. & Rozenblatt, S. (1988) *Mol. Biochem. Parasitol.* **31,** 27–34.
44. Edman, U., Meza, I. & Agabian, N. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3024–3028.
45. Brown, S. D., Huang, J. & van der Ploeg, L. H. T. (1992) *Mol. Cell. Biol.* **12,** 2644–2652.
46. Zomerdijk, J. C. B. M., Kieft, R., Shiels, P. G. & Borst, P. (1991) *Nucleic Acids Res.* **19,** 5153–5158.
47. Sherman, D. R., Janz, L., Hug, M. & Clayton, C. (1991) *EMBO J.* **10,** 3379–3386.
48. Lanzer, M., de Bruin, D. & Ravetch, J. V. (1992) *EMBO J.* **11,** 1949–1955.
49. Lanzer, M., de Bruin, D. & Ravetch, J. V. (1992) *Nucleic Acids Res.* **20,** 3051–3056.
50. Sogin, M. L. (1991) *Curr. Opin. Genet. Dev.* **1,** 457–463.